VP-Rec: A Hybrid Image Recommender Using Visual Perception Network

Crícia Z. Felício*‡, Claudianne M. M. de Almeida†, Guilherme Alves‡, Fabíola S. F. Pereira‡, Klérisson V. R. Paixão‡, Sandra de Amo‡, Celia A. Z. Barcelos‡
*Federal Institute of Triângulo Mineiro, IFTM
E-mail: cricia@iftm.edu.br
‡Federal University of Uberlândia, UFU
Uberlândia (MG), Brazil
E-mail:{guilhermealves, fabiola.pereira, klerisson, deamo, celiazb}@ufu.br
†Federal Institute of Norte de Minas Gerais, IFNMG
Arinos (MG), Brazil
E-mail: claudianne.almeida@ifnmg.edu.br

***Abstract*—A requirement for a great user experience is to meet the exact needs for the usage of a recommender system. Such systems need user's historical preferences to reasonably perform, which might not be the case for a cold-start user. This paper presents VP-Rec, a hybrid image recommender system that addresses the new user cold-start problem. VP-Rec combines user visual perception and pairwise preferences as source of information to perform recommendations. First, we infer pairwise preferences from users ratings. Next, we build visual perception networks linking users according to their visual attention similarities. From these two inferred structures, we build consensual prediction models, so that when a new user enters the system, we capture his visual attention and choose the best model that fits him. The system has been tested on two image datasets, getting important improvements in terms of ranking quality (nDCG) when applied to new user cold-start scenario against state-of-art recommender systems.**

***Index Terms*—Visual perception, Recommender system, User preferences, Eye tracking**

I. Introduction

Recommender systems (RS) are in our everyday life. We are usually asked to make choices without enough personal experience of the alternatives. So, we rely on others' recommendations and that is why RS have become ubiquitous nowadays. To do recommendations, those systems exploit users' previous choices and predict new products that would fulfill users' expectations. However, RS often face user cold-start problem [1], which is the challenge of recommending to users without preferences records. This lack of information leads RS to low accuracy levels and poor users experiences, that might affect the business performance.

Reliable user cold-start solutions do exist. The standard path is to infer implicit contextual information of the new user to work around cold-start problem. As contextual information we can mention social information [2], user click behavior [3], location-based information [4] and, more recently, user visual perception [5], [6]. In fact, tracking users eyes movements to capture their attention became an important source of knowledge with the accessibility to emerging technologies like smartphones cameras or eye tracking devices.

Melo et al. [5] proposed a content-based image recommendation approach applied to clothing shopping. Their approach uses items' ratings combined with users' visual attention. The goal is to recommend clothes similar to clothes already well rated by a user. Similarity among clothes is given by a measure calculated from visual attention similarity between them. Such approach achieves reasonable accuracy levels, but it does not deal with user cold-start problem.

In our prior work [6], we briefly introduce the idea of using visual attention to infer visual perception networks. The intuition is that users with similar visual perceptions have similar tastes. For instance, Figure 1 shows a painting containing two main scenes: a cat and a dog [1]. Some people looking at the painting might focus their attention to the cat. Others, to the dog. We can have two distinct groups of users. Thus, we explore users similarities within a single group to recommend items.

Fig. 1: Painting of a Dog and Cat. Some people might focus their attention to the cat, but others to the dog.

In this paper, we expand on these earlier works [5], [6] by combining *user visual perception* with prediction models of *pairwise preferences*. Pairwise preference is a specific type of opinion that establishes an order relation between two objects. For example, when a user says: "I prefer surrealism than cubism", we clearly identify his preference to paintings of the

[1] Oil Painting of a Dog and Cat, available at http://www.dailypainters.com/paintings/138359/Oil-Painting-of-a-Dog-and-Cat/Nancy-Spielman

TABLE I: Relational schema of paintings images.

	Title	Decade	Artist	Type	Art Movement
I_1	Dora Maar	1930	Picasso	Portrait	Surrealism
I_2	Portrait of Gala	1930	Dali	Portrait	Surrealism
I_3	Shades of Night	1930	Dali	Landscape	Surrealism
I_4	Nusch Eluard	1930	Picasso	Portrait	Cubism
I_5	Bust of a woman	1940	Picasso	Portrait	Cubism
I_6	Summer night	1920	Dali	Landscape	Surrealism
I_7	The Bleeding Roses	1930	Dali	Nudism	Surrealism
I_8	The Persistence of Memory	1930	Dali	Landscape	Surrealism

TABLE II: Users ratings over painting images.

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8
u_2	5	2	4	1	5	2	-	-
u_3	4	1	4	1	5	2	5	5
u_4	2	5	3	5	-	-	-	-
u_7	2	-	-	5	2	-	-	-
u_5	1	-	2	4	2	4	-	-
u_6	-	-	2	4	1	-	5	-

surrealism movement over cubism. PrefRec [7] is a hybrid recommender system that uses preferences to build prediction models. The advantage of recommending with preferences is that it does not suffer of: (i) lack of *Consistency*, which is incompatible comparison of users' ratings on same scale, for example, on 1 to 5 star ratings scale, a 4 rating from user X might be comparable to a 5 rating for user Y; (ii) lack of *Resolution*, this problem states that any numeric scale for ratings, say 1 to 5 stars, may be not capture all the users interests without loss of information [8].

Our new approach, called VP-Rec, uses visual perception to recommend images in a pairwise preference fashion. Therefore, it takes the advantages aforementioned, besides been a hybrid recommender systems. Instead of using only historical ratings, items features are applied to create the recommendation model and visual perception is used to define the items recommendation. The hypothesis is that *matching new people with existing people that present similar visual perceptions might help on providing accurate recommendations for cold-start users*. We address this by investigating three research questions:

RQ1: How effective is VP-Rec for cold-start user?

RQ2: How is the performance of VP-Rec under data sparsity?

RQ3: What is the performance comparison of matrix factorization approaches on users with observed ratings versus VP-Rec?

We compare our approach with four state-of-art social recommender system in terms of *nDCG* metric. Our results show that VP-Rec increases up to 90% the ranking quality compared to those systems.

II. Background

In this section we introduce the main concepts underlying VP-Rec. To enhance readability, we give an illustrative example along with the problem formalism. The focus is on how the prediction models are built and how the recommendation phase works.

Input and Output. Let $\mathcal{I} = \{I_1, ..., I_m\}$ be a set of images, and $\mathcal{U} = \{u_1, ..., u_n\}$ be a set of users. Let $RI(A_1, ..., A_p)$ be a relational scheme related to images, and $RU(A_{p+1}, ..., A_q)$ be a relational scheme related to users. The user-item rating matrix is represented by $\mathcal{R} = [r_{u,I}]_{m\times n}$, where each entry $r_{u,I}$ represents the rating given by user u on item image $I \in \mathcal{I}$.

Pairwise preference recommender systems predict the preference between a pair of items with missing values in the user-item rating matrix. On the other hand, in traditional recommender systems, the recommendation task is based on the predictions of the missing values in the user-item rating matrix. Both types of systems have the same output, a ranking of items where the k top-ranked are recommended.

Example. Table I shows an example of relational schema with attributes of 8 paintings images. A user-item rating matrix with the same 8 images and 6 users is exemplified in Table II.

PrefRec [7] is the hybrid approach we will extend with visual perception information. We focus in explain the PrefRec phases: (1) the Model Building, and (2) the Recommendation.

PrefRec Model Building Phase. In the first phase, PrefRec tasks include *Clustering user-item rating matrix* and *Preference Mining*. The goal is get a set of recommendation models to use in Recommendation Phase.

A) Clustering user-item rating matrix: PrefRec proposed to cluster users according to their preferences, using a distance function and a clustering algorithm. The preferences of each user u_t is represented by the row $\mathcal{R}_{u_t}$ of the user-item rating matrix $\mathcal{R}$. The output of the clustering algorithm is a set of clusters C^r, where each cluster C^r_j has a set of users with the most similar preferences (Pref-clusters). For each Pref-cluster C^r_j, a consensus operator is applied to compute V_j, the consensual preference vector of C^r_j. $V_{j,k}$ is the average rating for item k in cluster C^r_j.

Example. To illustrate these activities, an example of clustering and consensus calculus can be seen in Table III. We cluster the users from Table II in three Pref-clusters, and compute a consensual preference vector for each cluster using the group average rating per item.

B) Preference Mining: PrefRec relies on CPrefMiner [9] algorithm to build a contextual preference model as recommendation model. Having the consensual preference vector from each Pref-cluster, the system could establish the preference relation between pairs of images.

A preference miner algorithm builds a recommendation model for each group using item's features. The set of recommendation models is $M = \{M_1 = (V_1, Pm_1), \ldots, M_K = (V_K, Pm_K)\}$, where K is the number of Pref-clusters, V_j is the consensual preference vector, and Pm_j is the preference model extracted from V_j and the items attributes.

In this scenario, a recommendation model is a contextual preference model. Thus, each model Pm_j in M is designed as a *Bayesian Preference Network* (BPN) over $RI(A_1, ..., A_p)$. A BPN is a pair (G, φ) where G is a directed acyclic graph in which each node is an attribute, and edges represent attribute

TABLE III: Three Pref-clusters from user-item rating matrix in Table II.

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8
u_2	5	2	4	1	5	2	-	-
u_3	4	1	4	1	5	2	5	5
V_1	4.5	1.5	4	1.0	5.0	2.0	5.0	5.0
u_4	2	5	3	5	-	-	-	-
u_7	2	-	-	5	2	-	-	-
V_2	2.0	5.0	3.0	5.0	2.0	-	-	-
u_5	1	-	2	4	2	4	-	-
u_6	-	-	2	4	1	-	5	-
V_3	1.0	-	2.0	4.0	1.5	4	5	-

TABLE IV: V_1 pairwise preference relation

$(I_1 > I_3)$
$(I_3 > I_6)$
$(I_5 > I_6)$
$(I_6 > I_2)$
$(I_5 > I_3)$
$(I_2 > I_4)$
$(I_7 > I_6)$
$(I_7 > I_1)$
$(I_8 > I_1)$

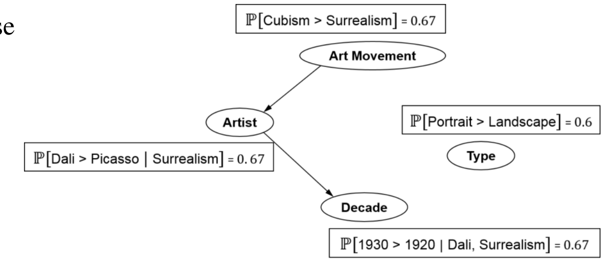


Fig. 2: Bayesian Preference Network **PNet**$_1$ over V_1 preferences.

dependency; φ is a mapping that associates to each node of G a set of conditional probabilities $\mathbb{P}[E_2|E_1]$ of the form of probability's rules: $A_1 = a_1 \wedge \ldots \wedge A_z = a_z \rightarrow B = b_1 > B = b_2$ where $A_1, \ldots, A_z$ and B are images attributes.

The constructing of a BPN is made in two steps: (1) the construction of a network structure represented by the graph G and (2) the computation of a set of parameters φ representing the conditional probabilities of the model. CPREFMINER [9] uses a genetic algorithm in the first phase to discover dependencies among attributes and then, compute conditional probabilities.

Example. To build the recommendation model for the first group in example aforementioned, PREFREC compute a preference relation over consensual preference vector V_1 as showed in Table IV. Then, the Bayesian preference network PNet$_1$ is computed (Fig. 6).

PREFREC Recommendation Phase. In recommendation phase, PREFREC needs previous ratings of a target user to choose an appropriate recommendation model. For a **new user** u_t, the algorithm computes the similarity between $\mathcal{R}_{u_t}$, row of user u_t in $\mathcal{R}$ matrix, and each consensual preference vector using a distance measure. Let V_j be the most similar consensual vector, then the recommendation model M_j is selected to make the pairwise predictions to user u_t. After that, the preference pairs are converted in a ranking.

Example: Suppose that V_1, depicted in Table III, is the most similar consensual vector for a **new user** u_t. Let us consider the BPN **PNet**$_1$ built over V_1 and depicted in Figure 2. This BPN allows to infer a preference ordering on items over relational schema *RI(Decade, Artist, Type, Art Movement)* of paintings images setting. For example, according to this ordering, painting I_5 = (1940, Picasso, Portrait, Cubism) is preferred than painting I_8 = (1930, Dali, Landscape, Surrealism). To conclude that, we execute the following steps:

1) We compute $\Delta(i_5, i_8)$, the set of attributes for which two paintings differ. Then, we remove attributes in $\Delta(i_5, i_8)$ that have at least one ancestor in the same set according to BPN structure and obtain min$(\Delta(i_5, i_8))$. In this example and considering **PNet**$_1$ structure, $\Delta(i_5, i_8)$ = {*Decade, Artist,Type,Art Movement*} and min$(\Delta(i_5, i_8))$ = {*Type,Art Movement*}.
2) Computing the probabilities: p_1 = *probability that* $i_5 > i_8 = \mathbb{P}[\textit{Portrait} > \textit{Landscape}] * \mathbb{P}[\textit{Cubism} > \textit{Surrealism}] = 0.6 * 0.67 = 0.402$; p_3 = *probability that* $i_8 > i_5 = \mathbb{P}[\textit{Landscape} > \textit{Portrait}] * \mathbb{P}[\textit{Surrealism} > \textit{Cubism}] = 0.4 * 0.33 = 0.132$; p_2 = *probability that* i_8 *and* i_5 *are incomparable* $= 1 - (p_1 + p_3) = 0.466$.

III. VP-REC APPROACH

To adopt visual perception as contextual information for recommendation systems, first, we rely on our *VP-Similarity Method* [6]. This method infers visual perception similarities among users. Then, we present our VP-REC Framework, which incorporates visual perception network on recommender systems.

A. VP-Similarity Method

Eye tracker devices capture information over user's visualization behavior (gaze positions, duration, sequence). We concentrate our definitions on gaze position and fixation length (length of time that visual attention lasts).

Definition 1 (Visual Fixation)**.** A visual fixation of a user u_t over an image $\mathcal{I}_k$ is a pair (p, f) where p is the position, represented by the pixels cluster centroid of that fixation, and f is the duration. We denominate $\mathcal{F}_{tk} = \{(p_1, f_1), ..., (p_z, f_z)\}$ the set of visual fixations of u_t over $\mathcal{I}_k$ (Fig. 3).

Definition 2 (Visual Perception)**.** Let the images in $\mathcal{I}$ be divided in r equal parts $Q = \{q_1, ..., q_r\}$ as illustrated in Fig. 4. From the positions and durations described in the set of visual fixations $\mathcal{F}_{tk}$, we call v_s the percentage of time that u_t fixed to $\mathcal{I}_k$ in each part q_s, for $1 \leq s \leq r$ (Fig. 5). The visual perception of a user u_t over an image $\mathcal{I}_k$ is defined as the vector $\mathcal{P}_{tk} = (v_1, ..., v_r)$. Finally, the visual perception of u_t over all images $\mathcal{I}$ is represented by the concatenation of all visual perceptions vectors from u_t: $\mathcal{P}_t = \mathcal{P}_{t1} \parallel ... \parallel \mathcal{P}_{tx}$. We denote by $\mathcal{P}$ the set of all users' visual perception vectors.

An example of visual perception can be seen in Table V. There are visual perceptions from 6 users over 2 images. Images are divided in 4 equal parts. For each user and each image, we have the percentage of time a given user fixed his visual attention in a corresponding part.

VP-similarity score is computed between two users u_1 and u_2 as the distance between their respective visual perceptions vectors $\mathcal{P}_1$ and $\mathcal{P}_2$. This distance is defined by the function

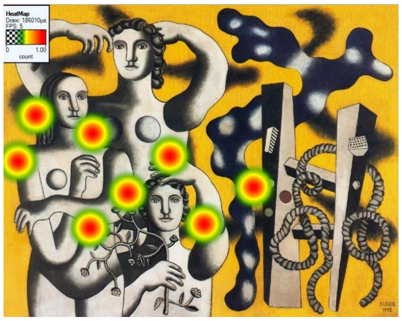


Fig. 3: Gaze positions and fixation length captured.

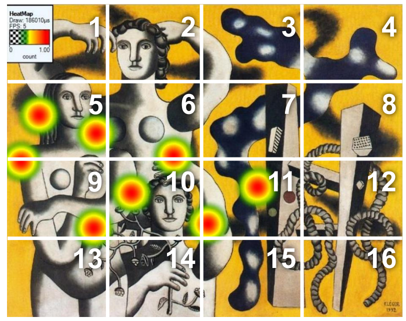


Fig. 4: Painting splits in sixteen equal parts.

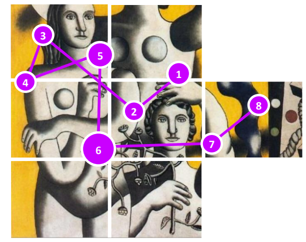


Fig. 5: Image parts with nonzero fixation length.

$l(u_1, u_2)$, where $l : \mathcal{P} \times \mathcal{P} \to \mathbb{R}$ and $l(u_1, u_2)$ can assume any classic similarity function like Euclidean distance, cosine similarity or Pearson distance correlation. By abuse of notation, we will write $l(u_1, u_2)$ as $l_{1,2}$. For example on Table V, we have that the VP-similarity score between u_4 and u_5, considering l as cosine similarity is 0.76 (* has been assumed as 0).

As we hypothesize that users with similar visual perceptions can be a good source for new user recommendation, we propose to cluster users according to their VP-similarity scores. In this paper, we use *K-means* as classical clustering algorithm, and refer to visual perception clusters as VP-clusters. This process is shown in the left side of Fig. 6.

We define as *cluster consensual vector* the vector containing the averages of all visual perceptions from users inside the same VP-cluster. Table V illustrates two VP-clusters and their respective consensual vectors $\hat{\mathcal{P}}_1$ and $\hat{\mathcal{P}}_2$. This notion is specially important on recommendation phase: when a target user u_t is added to the system, some visual perception of him is collected. Our VP-Similarity method generates the visual perception vector $\mathcal{P}_t$ of u_t, and a VP-similarity score between u_t and each VP-cluster C_j is computed. We denote $\delta_{t,k}$ as the VP-similarity score between a user u_t and a VP-cluster C_j (right side of Fig. 6). This notation is similar to l, previously defined. The goal is to find the most similar VP-cluster concerning the target user and associate him to the group. With the VP-clusters information the system will infer and update the Visual Perception Network and use it in the recommendation process (see Section III-B).

TABLE V: Users' visual perception over two images of paintings dataset.

	$\mathcal{I}_1$				$\mathcal{I}_2$			
	q_1	q_2	q_3	q_4	q_1	q_2	q_3	q_4
u_1	0.50	0.10	0.40	0.00	*	*	*	*
u_2	0.60	0.20	0.10	0.10	0.10	0.70	0.10	0.10
u_3	0.40	0.40	0.20	0.00	0.00	0.90	0.10	0.00
$\hat{\mathcal{P}}_1$	**0.50**	**0.23**	**0.23**	**0.03**	**0.05**	**0.80**	**0.10**	**0.05**
u_4	*	*	*	*	0.75	0.08	0.05	0.12
u_5	0.05	0.25	0.20	0.50	0.70	0.05	0.10	0.15
u_6	0.15	0.20	0.25	0.40	0.82	0.02	0.10	0.06
$\hat{\mathcal{P}}_2$	**0.10**	**0.23**	**0.23**	**0.45**	**0.76**	**0.05**	**0.08**	**0.11**

B. VP-Rec Framework

In this work, we propose an approach to incorporate VP-similarity in pairwise recommender systems to deal with cold-start problem. Figure 7 shows an overview of VP-REC framework.

Building Visual Perception Network: Given the users' visual perception over the set of images $\mathcal{I}$, the users can be clustered (as described in Section III-A) according to the visual perception (Module 1), generating a set of VP-clusters. Each VP-cluster C_j comprises a set of users and one consensual vector. Let $G = (V, E)$ be the visual perception network (VP-network) and u_t and u_v vertices of this graph. The VP-Network is build connecting all users in the same VP-cluster. Then, a set of neighbors of a user $u_t \in C_j$ is $N(u_t) = \{u_v | u_v \in V \wedge (u_t, u_v) \in E \wedge (u_v \in C_j)\}$.

Updating Visual Perception Network: Update in VP-Network have to be made when a user is added to a VP-cluster or a user is take out from one. When a user u_t is added to a VP-cluster C_j, we will insert edges on the VP-Network connecting u_t with each u_v in the same cluster. On the other hand, if a user u_t is take out from one VP-cluster C_j we will drop from the VP-Network all u_t's connections with users in C_j. These situations can happen when a new user is added to the system or an old user move to another cluster.

Building Recommendation Models: To build the recommendation models VP-REC, as PREF-REC does, computes the

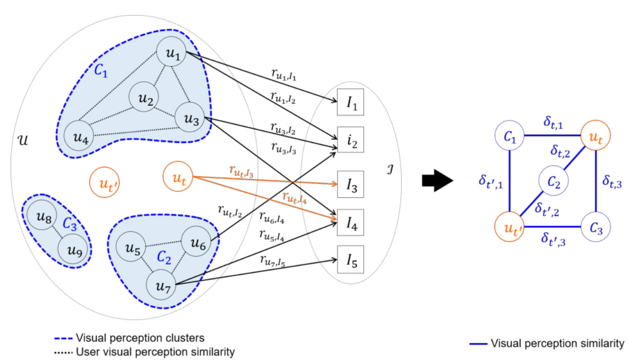


Fig. 6: Visual perception clusters and users' ratings (left), selection of visual perception cluster for $u_{t'}$ (cold start) and u_t (right).

clustering of $\mathcal{R}$ matrix and mining the preferences. Clustering the rows of user-item rating matrix $\mathcal{R}$ results in a set of Pref-clusters C^r. For each Pref-cluster C^r_j we apply a consensus operator to get a consensual preference vector V_j, where each position has the average ratings per item. From each V_j and the images features, we apply CPREFMINER algorithm [9] (Module 2) and has as output a preference model Pm_j. After building recommendations models we have a set of recommendation models $M_{vp} = \{M_{vp_0} = (C^r_1, V_1, Pm_1), \ldots, M_K = (C^r_K, V_K, Pm_K)\}$, where K is the number of Pref-clusters and each C^r_j represent the set of users in the Pref-cluster. Note that the set of users in a cluster was not used by PREFREC, but is necessary to VP-REC locate the recommendation models of the target user's neighbors.

VP-REC Recommendation: VP-REC method chooses between consensual recommendation models the most suitable for a new user. To recommend for a user is necessary to have visual perception information from him due the neighborhood is given by the VP-Network. In VP-REC, given a target user u_t and his neighbors $(N(u_t))$, the first task is select the recommendation model Pm_j corresponding to the Pref-cluster C^r_j with more visual perception neighbors. Pm_j is used to infer the preference between pairs of images in $\mathcal{I}$. We build a ranking using the set of predicted preferences between image pairs (Module 4) and evaluate the ranking quality over the top-k images.

Example: Consider a new user u_8 that is more similar, according to his visual perception, to VP-cluster C_2 (Table V). So, the set of u_8's neighbors is $N(u_8) = \{u_4, u_5, u_6\}$. At Table III we can see that u_4 is on Pref-cluster C^r_2 and u_5, u_6 is on C^r_3. How C^r_3 is the Pref-cluster with more neighbors, we will apply the recommendation model Pm_3 to make predictions to user u_8.

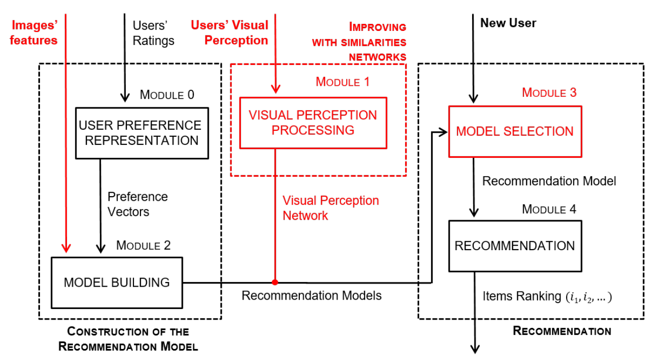


Fig. 7: VP-Rec Framework comprises four modules.

IV. EXPERIMENTAL SETUP

A. Dataset

There are several visual perception datasets, but for evaluating our recommendation model a suitable dataset must have item's attributes and ratings. Given the various factors that may influence recommendation systems, we analyze two different sets:

Paintings Dataset. We recruited 193 volunteers for rating 200 paintings, which were randomly chosen between 605 paintings public available at http://pintura.aut.org/. For each volunteer, an eye tracker device captures eye movements on each painting displayed on the 22' monitor with image resolution of 500 x 700 pixels. The paintings are composed by epoch, art movement, country, artist, type, color intensity and hue (image attributes). The volunteer should rate each painting in a 1-5 scale according to its preference.

Clothing Dataset. Melo et al. [5] also recruited volunteers to rate a clothing dataset. Hence, the full set is composed by two subsets of ratings over female and masculine clothing. In addition, they also collected visual attention through an eye tracker device. Clothing specific attributes are composed by class body, category, predominant color, color intensity, pattern, shape, size and sleeve. From original dataset we got only items rated in common among all users because we want to test networked information. Table VI summarizes datasets statistics.

TABLE VI: Paintings and Clothing dataset features.

Features	Paintings	Female-Clothing	Male-Clothing
# of users	194	121	120
# of items	605	210	210
# of ratings	38,753	25,396	25,193
Sparsity (%)	67.00	0.05	0.03
Links	28,992	7,204	9,531
Average # of ratings	199.88	209.88	209.94

B. Comparison Methods and Parameter Settings

To assess the effectiveness of VP-Rec, we compare it with four renowned recommenders:

PMF: A probabilistic matrix factorization approach [10]. This is the unique comparison method that does not use VP-similarity information. This method can be seen as a general baseline algorithm.

SoRec: A social recommender that uses probabilistic matrix factorization by employing both users' social network information and rating records [2]. This method is well recognized for the ability to deal with cold-start user, notably with full cold-star ones.

TrustMF: An adaption of matrix factorization technique to map users in terms of their trust relationship, aiming to reflect reciprocal users' influence on their own opinions [11]. Because this method showed remarkably results on dealing with cold-start users, we also select it to compare ours against to.

SocialMF: This method is a model-based matrix factorization approach that also explores the concept of trusting among users, but in the sense of propagation into the model [12]. This method was also tested against cold-start users.

Parameter Settings. VP-Similarity scores were computed splitting images in 4 equal parts. All methods make use of the visual perception generated by Module 1 of VP-Rec. We use LibRec [13] library implementation of SoRec, SocialMF, TrustMF and PMF methods with default parameters. For matrix factorization approaches the experiments were executed with 10 latent factors and number of interactions equal to 100.

VP-Rec cluster algorithm is K-means and the distance measure is Euclidean. We test several cluster size for preference and visual perception. Then for Pref-clusters the optimal numbers are 9 clusters for Painting dataset, 9 for Female-Clothing and 6 for Male-Clothing. To VP-clusters the optimal number is 2 clusters for all datasets.

C. Evaluation Protocols

We performed two classes of experiments reflecting differing numbers of ratings available to train each method. The first protocol, called **0-ratings protocol**, is basically the standard leave-one-out cross-validation, where the number of folds is equals to the number of instances in the dataset. Thus, each recommender system is applied once for each instance, using all other instances as a training set, but one selected as a single-user test.

We train the system with all users but one, which is the one selected for testing purpose. Note that none item ratings from the testing user is given to the system. Thus, we simulate a realistic cold-start scenario. In the second set of experiments, we apply the standard **five-fold cross-validation**.

With social approaches, we replace the required social network information by our visual perception network. Although our network is not a real social network, it is build based on the homophily assumption [14], which states that users linked with each other in social networks tend to have similar tastes, hence we linked users based on their visual perceptions similarities. Furthermore, we aim to investigate human visual attention to bootstrap recommender systems, mainly to handle cold-start problem. Because social recommenders is well known for dealing with new users, we chose them to compare to our approach.

V. Results and Discussions

Here, we assess the effectiveness of VP-Rec approach for item recommendation. In particular, we aim to answer our three research questions:

A. How effective is VP-Rec for cold-start user? (RQ1)

We assess the prediction quality of visual perception approaches among the state-of-art recommenders presented in Section IV-B. Table VII shows the result of this comparison in terms of ***nDCG*** rank size of 5, 10, 15, and 20 for items recommended in our three datasets (Paintings, Female-Clothing, and Male-Clothing).

The experimental results, for 0-ratings protocol, show the superiority of VP-Rec over all datasets. In particular, its performance might be explained because it needs none rating to build its recommendation model, which is the situation met in real applications. The recommendation for a 0-rating user u_k is then made selecting the consensual model according to u_k's visual perception network. Inside u_k VP-Network we can have distinct Pref-clusters, and VP-Rec chooses the one that contains more users. Recalling RQ1, this attests the effectiveness of apply visual perception for 0-ratings user in contrast to others social approaches.

We checked the normality and homogeneity of the nDCG results for each method using Shapiro and Bartlett test. We observed that the results values are not normally distributed and not homogeneous. Therefore, we performed the global comparisons with Kruskal-Wallis test. Our approach, with 95% of confidence, produced significant higher-quality results.

TABLE VII: nDCG for cold-start scenario (0-rating) against our three datasets.

(a) Paintings

Approach	Size of Rank			
	@5	@10	@15	@20
SoRec	0.8332 ± .126	0.8301 ± .110	0.8258 ± .101	0.8219 ± .098
SocialMF	0.8086 ± .123	0.8051 ± .103	0.8015 ± .097	0.8028 ± .091
TrustMF	0.6337 ± .145	0.6325 ± .127	0.6348 ± .122	0.6406 ± .118
PMF	0.6263 ± .157	0.6348 ± .135	0.6394 ± .128	0.6441 ± .118
VP-Rec	**0.9707** ± .053	**0.9616** ± .048	**0.9530** ± .101	**0.9457** ± .049

(b) Female-Clothing

Approach	Size of Rank			
	@5	@10	@15	@20
SoRec	0.7662 ± .157	0.7559 ± .137	0.7572± .128	0.7632 ± .119
SocialMF	0.7569 ± .155	0.7559 ± .135	0.7572 ± .127	0.7632 ± .122
TrustMF	0.6062 ± .139	0.6139 ± .122	0.6154 ± .118	0.6221 ± .113
PMF	0.5987 ± .162	0.5977 ± .134	0.6050 ± .122	0.6098 ± .114
VP-Rec	**0.9352**± .079	**0.9202** ± .078	**0.9107**± .073	**0.9130** ± .073

(c) Male-Clothing

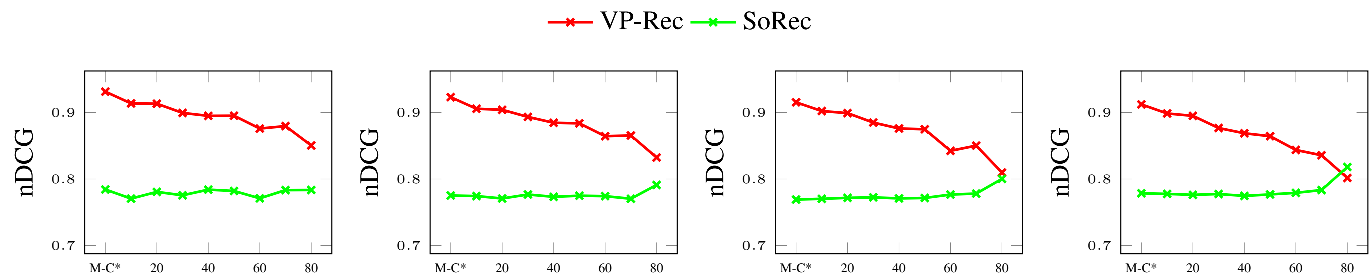
Approach	Size of Rank			
	@5	@10	@15	@20
SoRec	0.7842 ± .129	0.7752 ± .115	0.7691± .105	0.7785 ± .098
SocialMF	0.7708 ± .132	0.7655 ± .118	0.7645 ± .111	0.7698 ± .099
TrustMF	0.5941 ± .167	0.5955 ± .146	0.5993 ± .134	0.6045 ± .126
PMF	0.5759± .145	0.5794 ± .130	0.5852 ± .121	0.5919 ± .115
VP-Rec	**0.9314** ± .077	**0.9231** ± .069	**0.9154** ± .068	**0.9122** ± .067

B. How is the performance of VP-Rec under data sparsity? (RQ2)

Sparsity is the percent of empty ratings in user-item rating matrix. We investigate RQ2 using eight subsets obtained from Male-Clothing by eliminating a certain amount of ratings, see Table VIII. The reason for these experiments is the fact that sparsity is a big challenge faced by recommendation systems in general [15]. The idea is to simulate sparse scenarios where input datasets contains too many item to be rated and few items rated per user. For instance, Male-Clothing$_{80}$ was obtained by eliminating around 80% of the ratings in a stratified manner [16], so that we keep homogeneous subgroups of the original set.

TABLE VIII: Male-Clothing sparser subsets.

Male-Clothing (Dataset)	# of Ratings (Average)	Ratings per user (Average)	Sparsity (%)
10	22,720	189.33	9.84
20	20,186	168.21	19.89
30	17,672	147.26	29.87
40	15,150	126.25	39.88
50	12,617	105.14	49.93
60	10,116	84.3	59.85
70	7,579	63.15	69.92
80	5,083	42.35	79.82



(a) nDCG@5 (b) nDCG@10 (c) nDCG@15 (d) nDCG@20

Fig. 8: nDCG scores across Male-Clothing sparser subsets.

Because VP-Rec and SoRec were the methods that achieved better results under cold-start scenario, we choose them to test and compare their results under sparse subsets. Figure 8 shows the performance of each method per subset.

We note that VP-Rec is substantially affected by data sparsity. Its performance decreases as the data sparsity increases. On the hand, SoRec presents better results under sparser subset, enough to overcome VP-Rec performance against the most sparse scenario (80% of sparsity). Overall, the results suggest that VP-Rec effectiveness might be related with dataset density. However, its results was only surpassed for rank size of 20 items.

C. What is the performance comparison of matrix factorization approaches on users with observed ratings versus VP-REC*? (RQ3)*

The last experiment investigates the performance of VP-REC, with no ratings, against traditional approaches with certain amount of ratings. The idea is to analyze to what extent visual perception data suffice to offer accurate recommendation in the image data.

We test using 5-fold-cross validation technique, providing 20% of items ratings from each test user to bootstrap each matrix factorization system recommender. In these experiments we have PMF using 80% of ratings to build the target user recommendation model. SoRec, SocialMF and TrustMF combine 80% of ratings with visual perception information for the same task. On the other hand, VP-Rec select a consensual recommendation model using only visual perception information. All methods make predictions over the same 20% of ratings.

The overall result was the same under 0-rating protocol, see Table IX. Again, we performed Kruskal-Walis statistical test and it shows that VP-Rec is superior with 95% of confidence. Using only visual perception to select a consensual recommendation model, instead of build a personalized one, our approach is a good alternative to recommend images.

TABLE IX: nDCG for 5-fold-cross-validation protocol against our three datasets.

(a) Paintings

Approach	Size of Rank			
	@5	@10	@15	@20
SoRec	0.8287 ± .093	0.8210 ± .071	0.8181 ± .060	0.8132 ± .054
SocialMF	0.6713 ± .108	0.6766 ± .083	0.6791 ± .071	0.6804 ± .064
TrustMF	0.7389 ± .117	0.7360 ± .090	0.7334 ± .079	0.7314 ± .072
PMF	0.6292 ± .129	0.6281 ± .099	0.6258 ± .084	0.6247 ± .075
VP-Rec	**0.9284** ± .082	**0.9144** ± .080	**0.9029** ± .082	**0.8938** ± .083

(b) Female-Clothing

Approach	Size of Rank			
	@5	@10	@15	@20
SoRec	0.7367 ± .113	0.7316 ± .087	0.7298 ± .073	0.7322 ± .065
SocialMF	0.5785 ± .129	0.5719 ± .099	0.5529 ± .082	0.5511 ± .074
TrustMF	0.6710 ± .121	0.6636 ± .093	0.6616 ± .082	0.6626 ± .075
PMF	0.5688 ± .123	0.5689 ± .094	0.5706 ± .081	0.5747 ± .074
VP-Rec	**0.9044** ± .086	**0.8886** ± .079	**0.8741** ± .080	**0.8607** ± .080

(c) Male-Clothing

Approach	Size of Rank			
	@5	@10	@15	@20
SoRec	0.7300 ± .117	0.7253 ± .093	0.7282± .081	0.7321 ± .074
SocialMF	0.6121 ± .115	0.6086 ± .088	0.6023 ± .075	0.6049 ± .068
TrustMF	0.6527 ± .146	0.6538 ± .119	0.6590 ± .107	0.6660 ± .098
PMF	0.5491 ± .124	0.5548 ± .096	0.5611 ± .084	0.5676± .077
VP-Rec	**0.9118**± .093	**0.9008** ± .087	**0.8924** ± .084	**0.8844** ± .082

VI. RELATED WORK

VP-Rec draws together research on recommender systems with prior literature on cold-start problem and image recommendation.

Cold-Start Problem. It has already been several years of research on this topic. The dominant, near-universal trend, to alleviate such problem is to explore user's social information [17]. Our own work has followed this standard path [18], [19]. Remarkably, Ma et al. proposed the classic approaches, dubbed, SoReg [20] and SoRec [2], by incorporating the social network information into the PMF model [10]. Because SoRec is well renowned for dealing with cold-start user we compared our result against it. But, to be fair, in this paper we are interested in explore visual perception network. Although SoRec achieves high scores of nDCG, our networked information is not a social network. We argue that different contexts, such as online clothing shopping, might requires different contextual information, and that is because we are investigating visual perception networks. For instance, Macedo et al. reported on event recommendation problem [21]. They argue that events published in social networks are intrinsically cold-start, because they are typically short-lived. Thus, they proposed a hybrid recommendation approach that exploits several events' contextual information, whereas our approach is specially tailored to image recommendation.

Image Recommendation. A pioneer study of Xu et al. uses similarity based on visual perception to build recommendation models [22]. The experiments involved only five users, contrasting Google, YouTube and their proposal in search queries results. Umemoto et al. proposed to relate users' eye movements with information seeking. Then, they rank search results to emphasize relevant parts on a Web page [23]. The work [24] also used gaze positions of a user in conjunction with facial expressions as two types of implicit user feedback within the context of personalized web page recommendation. Those works did not handle images or videos' elements, just text content in search queries. Besides image recommendation being a thriving research field, another motivation is to complement the work of Melo et al. [5]. They proposed a content-based filtering enhanced by human visual attention applied to clothing recommendation. This approach is specific for clothes domain and relays on visual attention similarity combined with the measures conventionally used in content-based image recommendation systems. Furthermore, they work is limited by user cold-start problem.

Our work is innovative in the sense that we incorporate *visual perception data* as a contextual information for image recommender systems. We use a clustering-based filtering approach that infers a visual perception network, mainly to tackle new user cold-start problem.

VII. Conclusion

In this paper we introduced VP-Rec, an approach to handle user cold-start problem in image recommendation. We proposed to combine *user's visual perception*, as a valuable source of contextual information, with prediction models based on *pairwise preferences*. We thorough evaluated VP-Rec against two images dataset and showed that our approach beat state-of-art recommender systems that handle contextual networks, reaching up to 90% of ranking quality.

The ability to handle visual perception networks introduced by VP-Rec opens several avenues for future research. We will exploit other ways to measure visual similarities among users and apply filters during the recommendation phase according to a visual perception similarity score. We also intend to experiment other visual contexts domains such as online dating services.

Acknowledgments

C. Z. Felício would like to thank the Federal Institute of Triângulo Mineiro for study leave granted. We also thank the Brazilian Research Agencies CAPES, CNPq and FAPEMIG for supporting this work.

References

[1] J. Bobadilla, F. Ortega, A. Hernando, and J. Bernal, "A collaborative filtering approach to mitigate the new user cold start problem," *Knowledge-Based Systems*, vol. 26, pp. 225 – 238, 2012.

[2] H. Ma, H. Yang, M. R. Lyu, and I. King, "Sorec: Social recommendation using probabilistic matrix factorization," in *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, ser. CIKM '08. ACM, 2008, pp. 931–940.

[3] J. Liu, P. Dolan, and E. R. Pedersen, "Personalized news recommendation based on click behavior," in *Proceedings of the 15th International Conference on Intelligent User Interfaces*, 2010, pp. 31–40.

[4] C. Cheng, H. Yang, I. King, and M. R. Lyu, "Fused matrix factorization with geographical and social influence in location-based social networks," in *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[5] E. Viriato de Melo, E. A. Nogueira, and D. Guliato, "Content-based filtering enhanced by human visual attention applied to clothing recommendation," in *Tools with Artificial Intelligence (ICTAI), 2015 IEEE 27th International Conference on*, Nov 2015, pp. 644–651.

[6] C. Z. Felício, C. M. M. de Almeida, G. Alves, F. S. F. Pereira, K. V. R. Paixão, and S. de Amo, "Visual perception similarities to improve the quality of user cold start recommendations," in *Advances in Artificial Intelligence: 29th Canadian Conference on Artificial Intelligence*, 2016, pp. 96–101.

[7] S. de Amo and C. Goncalves, "Towards a tunable framework for recommendation systems based on pairwise preference mining algorithms," in *Proceedings of the 27th Canadian Conference on Artificial Intelligence*, 2014, pp. 282–288.

[8] S. Balakrishnan and S. Chopra, "Two of a kind or the ratings game? adaptive pairwise preferences and latent factor models," *Frontiers of Computer Science*, vol. 6, no. 2, pp. 197–208, 2012.

[9] S. de Amo, M. L. P. Bueno, G. Alves, and N. F. F. da Silva, "Mining user contextual preferences," *Journal of Information and Data Management*, vol. 4, no. 1, pp. 37–46, 2013.

[10] R. Salakhutdinov and A. Mnih, "Probabilistic matrix factorization," in *Advances in Neural Information Processing Systems*, vol. 20, 2008.

[11] B. Yang, Y. Lei, D. Liu, and J. Liu, "Social collaborative filtering by trust," in *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, ser. IJCAI. AAAI Press, 2013, pp. 2747–2753.

[12] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of the Fourth ACM Conference on Recommender Systems*, ser. RecSys '10. ACM, 2010, pp. 135–142.

[13] G. Guo, J. Zhang, Z. Sun, and N. Yorke-Smith, "Librec: A java library for recommender systems," in *23rd Conference on User Modeling, Adaptation, and Personalization (UMAP)*, 2015.

[14] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annual review of sociology*, pp. 415–444, 2001.

[15] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734–749, Jun. 2005.

[16] M. P. Cohen, *International Encyclopedia of Statistical Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, ch. Stratified Sampling, pp. 1547–1550.

[17] H. Kautz, B. Selman, and M. Shah, "Referral web: Combining social networks and collaborative filtering," *Commun. ACM*, vol. 40, no. 3, pp. 63–65, Mar. 1997.

[18] C. Z. Felício, K. V. R. Paixão, G. Alves, and S. de Amo, "Social prefrec framework:leveraging recommender systems based on social information," in *Proceedings of the 3rd Symposium on Knowledge Discovery, Mining and Learning*, 2015, pp. 66–73.

[19] C. Z. Felício, K. V. R. Paixão, G. Alves, S. de Amo, and P. Preux, "Exploiting social information in pairwise preference recommender system," *Journal of Information and Data Management*, (To appear).

[20] H. Ma, T. C. Zhou, M. R. Lyu, and I. King, "Improving recommender systems by incorporating social contextual information," *ACM Trans. Inf. Syst.*, vol. 29, no. 2, pp. 9:1–9:23, Apr. 2011.

[21] A. Q. Macedo, L. B. Marinho, and R. L. Santos, "Context-aware event recommendation in event-based social networks," in *Proceedings of the 9th ACM Conference on Recommender Systems*, ser. RecSys '15. New York, NY, USA: ACM, 2015, pp. 123–130.

[22] S. Xu, H. Jiang, and F. C. Lau, "Personalized online document, image and video recommendation via commodity eye-tracking," in *Proceedings of the 2008 ACM Conference on Recommender Systems*. New York, NY, USA: ACM, 2008, pp. 83–90.

[23] K. Umemoto, T. Yamamoto, S. Nakamura, and K. Tanaka, "Search intent estimation from user's eye movements for supporting information seeking," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2012, pp. 349–356.

[24] S. Xu, H. Jiang, and F. C. M. Lau, "Observing facial expressions and gaze positions for personalized webpage recommendation," in *Proceedings of the 12th International Conference on Electronic Commerce: Roadmap for the Future of Electronic Business*, 2010, pp. 78–87.